

Birden çok uygulamaya dayalı yöntemler alanyazında genellikle benzer ölçekler geçerliliği, hâlihazır geçerlilik, uygunluk geçerliliği olarak da geçer.

Birden çok uygulamaya dayalı yöntemlerle güvenilirlik kanıtı elde etmek biraz zahmetlidir. Test-tekrar test yönteminde testi farklı zamanda ikinci kez uygulamak için grup bulmak bazen zor olabilir. Ayrıca katılımcı kaybı yaşamak da olasıdır. Tek uygulamaya dayalı yöntemler tek bir test, tek bir grup ve tek bir uygulama gerektirmektedir. Dolayısıyla güvenilirlik kanıtı elde etmek daha pratiktir (Özellikle kararlılık katsayısının hesaplanmadığı durumlarda güvenilirliğin bir boyutu tabii ki eksik kalır.). Bir test bir gruba, bir kez uygulanıyorsa test kendi içinde değerlendirilecek demektir.

Güvenilirliğin bir boyutu da testin tutarlılığıdır. Tek uygulamaya dayalı güvenilirlik sorgulama yöntemleri ile "Test kendi içinde tutarlı bir bütün oluşturur mu?" sorusuna yanıt aranır. Bu nedenle bu yöntemlerin tümünden elde edilen katsayı **iç tutarlılık katsayısı** olarak adlandırılır.

### 2.5.2. TEST YARILAMA (EŞDEĞER YARILAR / İKİ YARI GÜVENİLİRLİĞİ) YÖNTEMİ:

Bu yöntemle güvenilirliği tahmin etmede uygulanmış bir test iki eşdeğer yarıya bölünür ve bireylerin iki yarıdan aldıkları puanlar arasındaki tutarlılık incelenir. Buradaki en temel sorunlardan biri testin iki eşdeğer yarıya nasıl bölünmesi gerektiği ile ilgilidir.

En sık başvurulan yöntemler: a) ilk yarı ve son yarı b) tek ve çift ve c) rastlantısaldır.

İlk ve son yarı yöntemi her test için uygun değildir.

Uygun olmadığı durumlar: Testteki

- maddeler basitten zora doğru sıralanmış ise
- maddeler konu içeriklerine ya da faktörlere göre kümelenerek yerleştirilmiş ise
- madde sayısı çok fazla ise

| | Türkçe Testi | $X_{1.YARI}$ | $X_{2.YARI}$ |
|---|---|---|---|
| Deniz | 48 | 20 | 28 |
| Şebnem | 98 | 50 | 48 |
| Ufuk | 66 | 32 | 34 |
| Işıl | 88 | 44 | 44 |
| Sinem | 32 | 20 | 12 |
| …… | 70 | 30 | 40 |
| …… | 62 | 28 | 34 |
| …… | 40 | 20 | 20 |
| …… | 60 | 28 | 32 |
| Sinan | 66 | 36 | 30 |

Testin yarısına ilişkin tutarlılığı verir. Bu korelasyon testin tamamına nasıl genellenir?

Korelasyon (r)

Spearman-Brown Eşitliği

İç tutarlılık katsayısı

*Şekil 11*: Test yarılama yöntemi çalışma örneği

Bu yöntemle güvenilirliği tahmin etmede, uygulanmış bir test iki eşdeğer yarıya bölünür ve bireylerin iki yarıdan aldıkları puanlar arasındaki tutarlılık incelenir. Şekil 11'de de görüldüğü üzere iki eşdeğer yarıdan alınan toplam puanlar arasında korelasyon hesaplanır ancak elde edilen katsayı korelasyonun yarı puanlar üzerinden elde edilmesi nedeniyle testin tamamına ilişkin bir tutarlılık bilgisi vermez, yarısına ilişkin verir. Diğer bir ifadeyle elde edilen katsayı eşdeğer yarılardan birinin güvenilirliği olarak kabul edilir. Testin tümüne ilişkin bir güvenilirlik katsayısı Spearman-Brown formülü aracılığıyla hesaplanır. Elde edilen katsayı testin tamamına ilişkin **iç tutarlılık** bağlamındaki güvenilirlik katsayısı olarak kabul edilir.



**2.5.3. KUDER-RICHARDSON 20 VE 21 YÖNTEMLERİ**: Kuder-Richarson 20 ve 21 ya da kısaca KR-20 ve KR-21, testin kendi içinde tutarlı bir bütün oluşturup oluşturmadığı ile ilgili bilgi verir. Bu nedenle her iki yöntemle de testin iç tutarlılığı değerlendirildiğinden bu yöntemlerden elde edilen katsayıya **iç tutarlılık katsayısı** adı verilir.

KR-20 ve KR-21 ile güvenilirlik kestirimi, puanlamanın kategorik olduğu ölçme araçları için uygundur. Puanlamanın kategorik olması, doğru yanıtlara 1 puan, yanlış ve boş bırakılan maddelere 0 puan vererek puanlamanın yapıldığı ya da maddenin yanıtının iki seçenekli verildiği durumlardır. Eğer testteki maddeler farklı ağırlıklarla puanlanmışsa ya da test puanları şans başarısı için düzeltilmişse bu formüller kullanılmaz.

**2.5.4. CRONBACH ALFA YÖNTEMİ**: Eğer ölçme aracının puanlaması çok kategorili ise KR-20 ile aynı mantık üzerine kurulu Cronbach alfa hesaplanır. Elde edilen katsayının ismi yine **iç tutarlılık katsayısı**dır. Derecelendirme ölçekleri puanlamanın çok kategorili olduğu araçlara örnek verilebilir.

**2.6. GÜVENİLİRLİĞİ VE GEÇERLİLİĞİ ARTIRMA YOLLARI**:

Güvenilirlik temelde tesadüfi hatalardan arınık olma durumu ise şüphesiz ki tesadüfi hata kaynaklarını azaltmaya yönelik eylemler güvenilirliği artıracaktır. Ancak bu noktada güvenilirliği artıran temel kavramlar ve eylemler konusunun ayrıntılarına girelim:

1. Bir testteki madde sayısı arttıkça birimler küçüldüğünden duyarlılık artar, bu nedenle hata miktarı azalacağından güvenilirlik artar.
2. Bir ölçme işleminde genel bir ilke olarak puanlayıcı sayısı arttıkça güvenilirlik artar. Birden çok kişinin yaptığı puanlamanın ortalamasını almak, bir kişinin verdiği puandan daha güvenilir olma eğilimindedir.
3. Puanlama nesnelliği, güvenilirliği büyük ölçüde etkiler. Bir testin farklı kişiler tarafından puanlanması ya da aynı kişinin farklı zamanlarda verdiği puanlar arasındaki tutarlılığa **puanlama güvenilirliği** adı verilir. Bir testten elde edilen puan, puanlayıcıya ya da zamana göre değişmiyorsa testin güvenilirliği artar. Elbette kullanılan ölçme aracının puanlamasının nesnel ve öznel olması bu noktada çok önemlidir. Çoktan seçmeli, derecelendirme ölçeği, kontrol listeleri gibi araçlar puanlaması nesnel araçlara örnek iken açık uçlu sorulardan oluşan araçlar, mülakatlar vb. ise puanlaması öznel araçlara örnek olarak verilebilir. Öznel araçların puanlama biçiminin mümkün olduğunca standartlaştırılarak nesnel hâle getirilmesi güvenilirlik açısından önemlidir.
4. KR-20 ve KR-21 yöntemlerinde bahsedildiği gibi testteki maddeler açısından benzeşiklik (homojenlik) arttıkça güvenilirliğin artması, diğer taraftan ayrışıklık (heterojenlik) arttıkça güvenilirliğin düşmesi beklenen bir durumdur.
5. Testten elde edilen puanların güvenilirlik kestirimi için veri elde edilecek grubun (örneklem) büyüklüğü arttıkça grubun heterojenleşmesinden dolayı güvenilirlik artar.
6. Aslında bir üst maddeyle de paralel bir biçimde maksimum performansı ölçen testlerde ortalama güçlüğe yaklaştıkça ($\bar{P} = .50$) grup heterojenleşir ve dolayısıyla güvenilirlik artar. Heterojen gruplarda testin ölçmeye konu olan özelliğe sahip olanla olmayanı ayırt etme gücünün artması beklenir. Grubun ölçülen özellik bağlamında heterojenleşmesi, büyük ölçüde örneklem sayısının artmasına ve eğer bilişsel bir özellik ölçülüyorsa testin ortalama güçlüğe yaklaşmasına bağlıdır.



7. Maddelerin dil bilgisi kurallarına uygun, açık ve anlaşılır yazılması belki de güvenilirliği en çok artıran etmenlerden biridir. Bu koşul sağlanmadığı takdirde diğer koşulların çok fazla önemi olmayabilir. İyi madde yazarı olmanın temel özelliklerinden biri de ana dilini iyi kullanma becerisidir. Dolayısıyla test yazar(lar)ının dil bilgisi ve imla kurallarına uygun bir biçimde maddeleri yazması, maddede yoklamak istediklerini yalın ve kısa bir biçimde ifade etmesi gerekir. İdealde bu konudaki temel ilke, bir maddenin tüm öğrenciler tarafından aynı biçimde anlaşılmasıdır.
8. Öğrencilere test uygulaması öncesinde ve gerekliyse süreçte yönerge vermek, test almaya güdülenme ve hazırbulunuşluluğun artması açısından önemlidir. Buna yönelik alınacak önlemler özellikle bireyden kaynaklanabilecek hataları asgariye indirir.
9. Fiziksel ortamın ses, sıcaklık, ışık, koku, dikkat dağıtıcı uyarıcılar, ergonomi vb. etmenler açısından uygun hâle getirilmesi yine güvenilirliği artıran bir başka etmendir. Fiziksel uyarıcılar açısından yanıtlama davranışını etkileyebilecek her türlü etmen güvenilirliği düşürecektir. Bu nedenle test uygulayıcılarının test ortamını düzenleme bağlamında sorumlulukları vardır. Özellikle zekâ, başarı gibi bilişsel özelliklerin ölçülmesinde bu etmenlerin önemi daha da artmaktadır.
10. Eğer süreli bir test uygulaması ise sürenin yeterli verilmesi gerekir. Eğer bir hız testi değilse tüm katılımcıların maddelerin tümüne ulaşması beklenen bir durumdur. Tabii ki hız testleri bu durumdan bağımsız olarak değerlendirilebilir.
11. Uygulamada bireylerin dikkatini dağıtacak ve/veya kaygısını artıracak davranışlardan kaçınılması, puanlama ya da veri girişi yapılırken dikkatli ve titiz davranılması ölçme işlemini yapan kişiden kaynaklanabilecek hataları azaltacağından dolayı güvenilirliği artırır.
12. Maddelerin teste düzgün yerleştirilmesi, okumayı güçleştirecek bir unsurun bulunmaması; test katılımcısının yaşına, gelişim düzeyine uygun bir punto büyüklüğü kullanılması; baskı hataları bulunmaması vb. etmenler güvenilirliği artırır.

Güvenilirliği artıran faktörler geçerliliği de artırır. Ancak geçerlilik yalnızca tesadüfi hatalardan değil, aynı zamanda sabit ve sistematik hatalardan da etkilenir. O nedenle öğretmenin tesadüfi hataların yanı sıra sabit ve sistematik hatalardan arındırmaya yönelik alacağı tüm tedbirler geçerliliği de artıracaktır.